AD-A051 860 INDIANA UNIV BLOOMINGTON DEPT OF MATHEMATICS F/G 12/1
LIMIT THEOREMS RELATED TO DISTRIBUTIONS OF CERTAIN LINEAR RANK --ETC(U)
FEB 78 M L PURI, N S RAJARAM AFOSR-76-2927
UNCLASSIFIED AFOSR-TR-78-0433 NL

| OF |

AD
A051 860

END
DATE
FILMED
4-78
DDC

Limit Theorems Related to Distributions of Certain Linear Rank Statistics with Stochastic Predictors

by

Madan L. Puri and Navaratna S. Rajaram
Indiana University and Kent State University

1. Introduction. Consider sequences of independent random variables $\{X_{ni}: 1 \leq i \leq n, n \geq 1\}$ with continuous cumulative distribution functions (cdf). A broad range of statistical tests based on ranks lead to the study of simple linear rank statistics of the type

$$S_n = \sum_{i=1}^{n} C_{ni}\, a_n(R_{ni}) \tag{1.1}$$

where R_{ni} is the rank of X_{ni} among $(X_{n1}, X_{n2}, \cdots, X_{nn})$; $\{C_{ni}: 1 \leq i \leq n\}$ and $\{a_n(i): 1 \leq i \leq n\}$ are sets of known constants. One usually assumes $a_n(i)$ (scores) to be generated by a known function φ: $(0, 1) \longrightarrow R$ in either of the following ways:

$$a_n(i) = \varphi(i/(n+1)), \tag{1.2}$$

$$a_n(i) = E\, \varphi(U_n^{(i)}), \quad 1 \leq i \leq n \tag{1.3}$$

where $U_n^{(i)}$ is the i-th order statistic in a random sample of size n from a uniform distribution over (0, 1).

Hájek (1968) obtained asymptotic normality of the statistic S_n defined in (1.1) with suitable conditions on C_{ni} and the score generating function φ.

AMS (1970) subject classification: 62G99.
Work supported by the Air Force Office of Scientific Research, AFSC, USAF under Grant No. 76-2927. Reproduction in part or in whole permitted for any purpose of the United States Government.

However, there exist situations where the C_{ni} can no longer be regarded as constants, being themselves random variables, known as "stochastic predictors." Corresponding to such stochastic predictors (and the sequence $\{X_{ni}\}$), we define a stochastic linear rank statistic by

$$(1.4) \qquad T_n = n^{-1/2} \sum_{i=1}^{n} C_{ni} \, a_n(R_{ni})$$

wherein all the quantities have already been defined.

This paper is devoted to the study of asymptotic distribution theory of the statistic T_n. Thus results of Hájek (1968) are a special case of our results for T_n. However, as opposed to Hájek, we obtain a family of limit laws which are weighted averages of the normal and degenerate to the normal in extreme cases. Thus the results are also of independent interest. In view of the rather complex nature of the conditionings involved in our analysis, it seems rather doubtful that Hájek's projection technique can be used in a situation as general as ours.

Although we take C_{ni} to be stochastic, it is clear that they cannot be entirely arbitrary. The restrictions which we shall impose on them are of two types. Firstly, they are concerned with their magnitudes and growth rates of certain fractional moments, stated explicitly in the theorems. Secondly, they are concerned with the measure theory and dependence among themselves and with the X_{ni}. Clearly, there is considerable latitude in the choice of the dependence structure and one can weaken one set of conditions by strengthening others. We shall, however, assume only that the X_{ni} are independent for a given set $\{C_{ni}: \ 1 \leq i \leq n\} = (\underset{\sim}{C}_n)$. (Conditional independence.) This situation is realized when, for example, the sequence $\{(X_{ni}, C_{ni})\}$ is composed of independent vectors. Even when they are exchangeable, it can be expressed in present framework.

We denote by F_{ni} the cdf of X_{ni} given $\underset{\sim}{C}_n$. (It is to be noted that F_{ni} can depend upon $\underset{\sim}{C}_n$ and is, therefore, a random variable, but F_{ni} is a.s. a distribution function.)

In view of the rather delicate nature of the conditioning needed in our proofs, it is desirable to set down the measure theoretic framework involved in our study. All measures considered are, of course, probability measures. These considerations are rather crucial in the proof of Proposition 3.2.

Let $(\Omega, \mathcal{A}, P)$ be a measure space rich enough to admit families of random variables $\underset{\sim}{C}_n$ and $\underset{\sim}{X}_n$, for $n \geq 1$. Let $(\Omega_i, \mathcal{B}_i)$, $i = 1, \cdots, n$ be measure spaces with respect to which $(\underset{\sim}{X}_i, \underset{\sim}{C}_i)$, $i = 1, \cdots, n$ are measurable. The product measurable space $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{B}_1 \times \cdots \times \mathcal{B}_n)$ is the space of points $(\omega_1, \cdots, \omega_n)$ together with the minimal σ-field over measurable rectangles $B_1 \times \cdots \times B_n$, where $B_i \in \mathcal{B}_i$, $i = 1, \cdots, n$. Then the product measurable space $(\Pi\Omega_n, \Pi\mathcal{B}_n)$ is the space of points $(\omega_1, \omega_2, \cdots)$, $\omega_i \in \Omega_i$, and the minimal σ-field over the measurable cylinders $(B_1 \times \cdots \times B_n) \times \prod_{k=n+1}^{\infty} \Omega_k$, $n = 1, 2, \cdots$ denoted by $\mathcal{A}_0$. Clearly $\mathcal{A}_0 \subset \mathcal{A}$.

Since we shall be conditioning with respect to $\underset{\sim}{C}_n$, we assume that the conditional probabilities for the $\underset{\sim}{X}_n$ given $\underset{\sim}{C}_n$ are all measurable with respect to $\mathcal{B}_n$. We also assume that the family $\{\underset{\sim}{C}_n\}$ determines a family $\{Q_n\}$ of regular conditional probabilities. That is to say, for each fixed $\omega_i \in \Omega_i$, $i = 1, \cdots, n-1$, $B_n \in \prod^n \mathcal{B}_k$, $Q_n(\omega_1, \cdots, \omega_{n-1}; B_n)$ is a.s.(P) a measure on $\prod^n \mathcal{B}_k$; and for a fixed B_n, a measurable function in $\omega_1, \cdots, \omega_{n-1}$, for all n.

2. Theorems. We now turn to the statements of our theorems. Denote

$$(2.1) \qquad H(x) = n^{-1} \sum_{i=1}^{n} F_{ni}(x)$$

$$(2.2) \qquad C(x) = n^{-1/2} \sum_{i=1}^{n} c_{ni} F_{ni}(x), \quad \mu_n = \int_0^1 \varphi(H(x))dC(x).$$

(It is to be noted that $H(x)$ may depend upon $\underset{\sim}{c}_n$.)

Theorem 2.1 (Representation Theorem). Let the scores $a_n(i)$, $1 \leq i \leq n$ be given by (1.2) and let the following conditions be satisfied.

$$(2.3) \qquad \max_{1\leq i\leq n} |c_{ni}| = O_p(1)$$

$$(2.4) \qquad |\varphi^{(i)}(t)| \leq K\{t(1-t)\}^{\delta-i-1/2}, \quad 0 < t < 1; \quad i = 0, 1$$ for some $K > 0$ and some $0 < \delta < 1/2$.

Then

$$(2.5) \qquad T_n = \mu_n + n^{-1/2} \sum_{i=1}^{n} \psi(X_{ni}, \underset{\sim}{C}_n) + D_n$$

where, for a given $\underset{\sim}{C}_n$,

$$(2.6) \qquad D_n \longrightarrow 0$$ in probability (conditionally) as $n \to \infty$.

$$(2.7) \qquad \psi(X_{ni}, \underset{\sim}{C}_n) = n^{-1} \sum_{j=1}^{n} (c_{nj} - c_{ni}) \int_{-\infty}^{\infty} \{I_{[x\leq y]} - F_{ni}(y)\}\varphi'(H(y))\, d\, F_{nj}(y)$$

and

$$(2.8) \qquad I_{[x\leq y]} = 1$$ if $x \leq y$, and 0 otherwise.

Furthermore, for a given $\underset{\sim}{C}_n$, the conditional distribution of $T_n - \mu_n$ is asymptotically normal, provided the corresponding value of the conditional variance is finite and nonzero.

Theorem 2.2 (Main Limit Theorem). In addition to the hypothesis of Theorem 2.1, let the following conditions be satisfied.

There exists an $\alpha \leq 1/2$, $0 < \alpha < \alpha(\delta) = 2\delta(1 - 2\delta)^{-1}$, $0 < \delta < 1/2$ such that

$$(2.9) \qquad E\{\max_{1 \leq i \leq n} |C_{ni}|^{2(1+\alpha)}\} = o(n^{\alpha});$$

and the family $\{\underset{\sim}{C}_n\}$ determines a family $\{Q_n\}$ of regular conditional probability measures in the sense described in section 1. We then have in the representation (2.5),

(2.10) $D_n \to 0$ in probability (unconditionally) as $n \to +\infty$ and

(2.11) T_n has a limit distribution if and only if the joint distribution of μ_n and the conditional variance $\sigma_n^2(\underset{\sim}{C}_n)$, say $W_n(\cdot, \cdot)$ converge weakly to a proper distribution function $W(\cdot, \cdot)$.

Furthermore, this limit law is weighted normal Φ_W whose characteristic function is given by

$$(2.12) \qquad f_W(u) = \int \exp(iux)d\,\Phi_W(x) = \int \exp(iu\mu - \frac{u^2\sigma^2}{2})dW(\mu, \sigma^2).$$

It must be noted that in the joint distribution of $(\mu_n, \sigma_n^2(\underset{\sim}{C}_n))$, it may be necessary to normalize μ_n by a sequence γ_n of constants (say $\gamma_n = E\,\mu_n$). Since this does not affect our analysis (we could consider $T_n - \gamma_n$), we shall disregard it.

Theorem 2.3 (Normal Convergence). Under the conditions of Theorem 2.2, the limit law of T_n is normal if and only if the law $W(\cdot, \cdot)$ is degenerate at some (μ_0, σ_0^2), $\sigma_0^2 \neq 0$.

It is obvious that in general the form of the limit law given by (2.12) can be quite complicated. It is possible to get a clearer idea of the kind of distributions which arise by looking at $T_n - \mu_n$. (Note that in the unconditional case μ_n is stochastic.) With this point in view, we give the following corollary and an example.

Corollary 2.1. *Under the conditions of Theorem 2.2, $T_n - \mu_n$ has a limit law if and only if the conditional variance $\sigma_n^2(\underset{\sim}{C}_n)$ converges in distribution to a proper random variable with distribution $W(\cdot)$. The limit law is then symmetric weighted normal Φ_W, (say) with the characteristic function*

$$g_W(u) = \int_0^\infty \exp(-1/2\, u^2\sigma^2)\, dW(\sigma^2). \tag{2.13}$$

Example. In Corollary 2.1, let σ^2 have an exponential distribution. Then

$$g_W(u) = \int_0^\infty \exp\left(-\frac{u^2\sigma^2}{2}\right) e^{-\sigma^2}\, d\sigma^2 = \left(1 + \frac{u^2}{2}\right)^{-1}.$$

Inversion of g_W by the Fourier-Stieltjes inversion formula yields the distribution Φ_W with the density

$$\frac{d}{dx}\Phi_W(x) = 2^{-1/2} \exp(-2^{1/2}|x|), \ -\infty < x < +\infty. \tag{2.14}$$

Thus, the form of Φ_W can differ quite drastically from the normal.

Remark. We note a "nonuniqueness" in our main results. (Theorem 2.2; Corollary 2.1.) In contradistinction to most of the limit theorems in statistical theory, we do not have a single limit distribution but a family of limit laws "parametrized" by the distributions W( ). It would be interesting and valuable to know what all distributions can arise. Equivalently, what distributions can be characterized at weighted normal.

Proofs of these results are given in Section 3. Similar results hold when $a_n(i)$ are given by (1.3). Our conditions appear to be broad enough to cover most practical situations. But the condition (2.9) can probably be weakened.

3. Proofs of the Theorems: We introduce the following processes:

$$(3.1) \qquad C_n(x) = n^{-1/2} \sum_{i=1}^{n} C_{ni} \, I_{[X_{ni} \leq x]}, \quad C(x) = E'[C_n(x)]$$

$$(3.2) \qquad H_n(x) = n^{-1} \sum_{i=1}^{n} I_{[X_{ni} \leq x]}, \quad H(x) = E'[H_n(x)]$$

$E'[\cdot]$ denotes $E[(\cdot)|\underset{\sim}{C}_n]$.

The following stochastic inequalities are obvious:

$$(3.3) \qquad |C_n(x)| \leq n^{1/2} \max_{1 \leq i \leq n} |C_{ni}| H_n(x)$$

and

$$(3.4) \qquad |C(x)| \leq n^{1/2} \max_{1 \leq i \leq n} |C_{ni}| H(x) \quad \text{a.s.}$$

Since the conditioning is with respect to $\underset{\sim}{C}_n$, inequalities and bounds hold a.s. Also note that $H(x)$ and $F_{ni}(x)$ are a.s. distribution functions. For the sake of brevity, we shall suppress the notation a.s.

We regard $C_n(x)$ as a random measure which assigns a weight C_{ni} to the point $x = X_{ni}$.

Proof of Theorem 2.1. From (1.2), (1.3) and (3.3), it follows that

$$(3.5) \qquad T_n = n^{-1/2} \sum_{i=1}^{n} C_{ni} \, \varphi\left(\frac{R_{ni}}{n+1}\right) = \int_{-\infty}^{\infty} \varphi\left(\frac{n H_n(x)}{n+1}\right) d\, C_n(x)$$

$$= \mu_n + B_{1n} + B_{2n} + \sum_{i=1}^{3} D_{in}$$

where μ_n is given by (2.2),

$$(3.6) \qquad B_{1n} = \int_{-\infty}^{\infty} \varphi(H(x))\, d[C_n(x) - C(x)]$$

$$(3.7) \qquad B_{2n} = \int_{-\infty}^{\infty} [H_n(x) - H(x)] \varphi'[H(x)]\, dC(x)$$

$$(3.8) \qquad D_{1n} = \frac{-1}{n+1} \int_{-\infty}^{\infty} H_n(x)\ \varphi'[H(x)]dC_n(x)$$

$$(3.9) \qquad D_{2n} = \int_{-\infty}^{\infty} [H_n(x) - H(x)]\varphi'\ [H(x)]\ d[C_n(x) - C(x)]$$

$$(3.10) \qquad D_{3n} = \int_{-\infty}^{\infty} \{\varphi[\frac{n}{n+1} H_n(x)] - \varphi[H(x)] - (\frac{n\ H_n(x)}{n+1} - H(x))\varphi'[H(x)]\}dC_n(x).$$

We shall first show that μ_n is well defined.

$$|\mu_n| = |\sum_{i=1}^{n} \frac{C_{ni}}{\sqrt{n}} \int_{-\infty}^{\infty} \varphi[H(x)]d\ F_{ni}(x)| \leq n^{1/2} \max_{1\leq i\leq n} |C_{ni}| \int_{-\infty}^{\infty} |\varphi[H(x)]|dH(x)$$

Since the integral is finite (by (2.4)), μ_n is a proper random variable.

Integrating B_{2n} by parts, we obtain

$$(3.11) \qquad B_{2n} = [H_n(x) - H(x)]B^*(x)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} B^*(x)d[H_n(x) - H(x)]$$

where

$B^*(x) = \int_{x_0}^{x} \varphi'[H(y)]dC(y)$, x_0 chosen arbitrarily such that $H(x_0) \neq 0$.

Let $\beta(x) = [H_n(x) - H(x)]B^*(x)$. Then

$$|\beta(x)| = n^{1/2}|H_n(x) - H(x)|n^{-1/2}| \int_{x_0}^{x} \varphi'[H(y)]dC(y)|$$

$$\leq O_p(1)\ n^{1/2}|H_n(x) - H(x)|\ |\int_{x_0}^{x} \varphi'[H(y)]dH(y)|,$$

$$\leq O_p(1)n^{1/2}|H_n(x) - H(x)|\{H(x)(1 - H(x))\}^{\delta-1/2}\cdot K.$$

Now by Puri-Sen (1971), given ϵ, $\delta' > 0$, there exists a constant $C(\epsilon, \delta')$, such that

$$P(\sup_x \frac{n^{1/2}|H_n(x) - H(x)|}{\{H(x)(1-H(x))\}^{1/2-\delta'}} > C(\epsilon, \delta')) < \epsilon.$$

Choosing $\delta' < \delta$, it follows that

$$|\beta(x)| \leq K \cdot O_p(1) \{H(x)[1 - H(x)]\}^{\delta-\delta'} C(\epsilon, \delta') \to 0$$

as $x \to \pm \infty$ for a fixed $\underset{\sim}{c}_n$. Unconditional convergence follows from Proposition 3.2 given a little later. Thus from (3.6), (3.7), and (3.11), we note that

$$B_{1n} + B_{2n} = \int_{-\infty}^{\infty} \varphi[H(x)]d[C_n(x) - C(x)] + \int_{-\infty}^{\infty} \varphi'[H(x)][H_n(x) - H(x)]dC(x).$$

Integrating B_{1n} by parts and collecting the coefficients of $[I_{[X_{ni} \leq \cdot]} - F_{ni}(\cdot)]$, we obtain the desired representation (2.5). Let now $\underset{\sim}{C}_n = \underset{\sim}{C}'_n$ be given. Then by Puri-Rajaram (1977), it follows that $\sum_{i=1}^{3} D_{in} \to 0$ in probability, provided $\max_{1 \leq i \leq n} |C_{ni}|$ is bounded, which by assumption (2.3) remains so with probability close to 1. The unconditional convergence is established in Proposition 3.2. To prove conditional normality, we verify the Liapunov condition. For simplicity, let us denote

$$(3.12) \qquad Y_{ni} = n^{-1/2} \psi(X_{ni}, \underset{\sim}{C}_n).$$

For conditional normality, it suffices to prove

$$(3.13) \qquad \lim_{n \to \infty} \sum_{i=1}^{n} E'|Y_{ni}|^{2(1+\alpha)} = 0 \text{ for some } \alpha > 0.$$

Denote

$$(3.14) \qquad E(\cdot \,|\underset{\sim}{C}_n) = E'(\cdot).$$

We obtain more than (3.13) in the following lemma. It will be needed to obtain the unconditional distribution of T_n.

Lemma 3.1. Under the assumptions of Theorem 2.1,

$$(3.15)\qquad \sum_{i=1}^{n} E'|Y_{ni}|^{2(1+\alpha)} = O(n^{-\alpha}) \max_{1\le i\le n} |C_{ni}|^{2(1+\alpha)} \quad \text{a.s.}$$

for each α, $0 < \alpha < \alpha(\delta) = 2\delta(1 - 2\delta)^{-1}$, where δ is given in (2.4).

Proof. We shall estimate the orders of magnitude for the terms of B_{1n} and B_{2n} separately, and then use the C_r-inequality. Recall that

$$(3.16)\qquad B_{2n} = \int_{-\infty}^{\infty} B^*(x)d[H_n(x) - H(x)] = n^{-1} \sum_{i=1}^{n} \{B^*(X_{ni}) - E'B^*(X_{ni})\}.$$

Take $0 < \alpha < \alpha(\delta)$. Then $2(1 + \alpha)(\delta - 1/2) > -1$. Next

$$(3.17)\qquad \sum_{i=1}^{n} E'[n^{-1}B^*(X_{ni})]^{2(1+\alpha)}$$

$$\le n^{-2(1+\alpha)} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left| \int_{x_0}^{x} \varphi'[H(y)]dC(y)\right|^{2(1+\alpha)} d\,F_{ni}(y)$$

$$\le n^{-2(1+\alpha)} \sum_{i=1}^{n} \{n^{1/2} \max_{1\le i\le n} |C_{ni}|\}^{2(1+\alpha)} \int_{-\infty}^{\infty} \left| \int_{x_0}^{x} \varphi'[H(y)]dH(y)\right|^{2(1+\alpha)} dF_{ni}(y)$$

$$\le \max_{1\le i\le n} |C_{ni}|^{2(1+\alpha)} n^{-(1+\alpha)} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{|\varphi[H(x)]| + |\varphi[H(x_0)]|\}^{2(1+\alpha)} dF_{ni}(y)$$

$$= \max_{1\le i\le n} |C_{ni}|^{2(1+\alpha)} n^{-\alpha} \int_{-\infty}^{\infty} \{|\varphi[H(x)]| + |\varphi[H(x_0)]|\}^{2(1+\alpha)} dH(x)$$

$$\le n^{-\alpha} K \max_{1\le i\le n} |C_{ni}|^{2(1+\alpha)} \int_{-\infty}^{\infty} \{H(x)(1-H(x))\}^{2(1+\alpha)(\delta-1/2)} dH(x)$$

$$= O(n^{-\alpha}) \max_{1\le i\le n} |C_{ni}|^{2(1+\alpha)},$$ since the integral is finite and

$\varphi[H(x_0)]$ is constant.

Since $E'|B^*(X_{ni})|^{2(1+\alpha)} \geq |E'B^*(X_{ni})|^{2(1+\alpha)}$ a.s. we can use the C_r-inequality for (3.16). Similar arguments hold for B_{1n}, and an application of the C_r-inequality yields the estimate (3.15). The Lemma is proved.

Remark: Since $\max_{1\leq i\leq n} |C_{ni}|$ is $O_p(1)$, it is obvious that the (conditional) asymptotic normality holds for each uniformly bounded set of values $\underset{\sim}{C}_n$, provided the corresponding value of the conditional variance converges to a finite, nonvanishing constant.

Proof of Theorem 2.2. We shall obtain a generalization of Theorem A, page 381, Loéve (1963). Let

$$\sigma^2_{ni}(\underset{\sim}{C}_n) = \mathrm{Var}(Y_{ni}|\underset{\sim}{C}_n)$$

and

$$U_{ni} = \psi(X_{ni}, \underset{\sim}{C}_n) + \mu_{ni};\ \mu_{ni} = n^{-1/2}C_{ni}\int_{-\infty}^{\infty} \varphi(H(x)\ d\ F_{ni}(x). \tag{3.18}$$

It is easily verified that

$$T_n = \sum_{i=1}^{n} U_{ni} + D_n = U_n + D_n, \text{ say.} \tag{3.19}$$

To obtain the asymptotic, unconditional distribution of U_n, we use the comparison theorem on pages 375-376 of Loéve (1963). Choose random variables V_{ni}, conditionally normal, with

$$E'(V_{ni}) = \mu_{ni},\ \mathrm{Var}'(V_{ni}) = \sigma^2_{ni}(\underset{\sim}{C}_n).$$

Letting

$$V_n = \sum_{i=1}^{n} V_{ni}$$

we shall compare V_n and U_n, taking $S = R$, $\ell = 0$, $m = 2$.

For $j \leq m$, we have $E'U^j_{ni} = E'\ V^j_{ni}$. For higher moments of V_{ni}, we shall bound them by the moments of V_{ni} as follows.

Since V_{ni} are conditionally normal with parameters μ_{ni} and $\sigma^2_{ni}(\underset{\sim}{C}_n)$, given $\gamma > 0$,

$$E'|V_{ni}|^{2+\gamma} \leq 2^{1+\gamma} (E'|V_{ni} - \mu_{ni}|^{2+\gamma} + |\mu_{ni}|^{2+\gamma}) \quad \text{a.s.}$$

It is easy to check (by normality)

$$E'|V_{ni} - \mu_{ni}|^{2+\gamma} = \{\sigma_{ni}(\underset{\sim}{C}_n)\}^{2+\gamma} \quad \text{a.s.}$$

Consequently, we have

$$\text{(3.20)} \qquad E'|V_{ni}|^{2+\gamma} \leq C\, E'|U_{ni}|^{2+\gamma} \quad \text{a.s.}$$

Next, in the comparison theorem, since $\ell = 0$, $S = R$, we need only verify conditions (iii and (iv), (Loéve, p. 376).

Note that $\int x^j \, dp'_{ni}(x) = \int x^j \, dq'_{ni}(x)$, $j = 1, 2$ where p'_{ni} is the conditional distribution of U_{ni} and q'_{ni} that V_{ni}. Hence

$$\text{(3.21)} \qquad \sum_{i=1}^{n} E \left| \int x^j (dp'_{ni}(x) - dq'_{ni}(x)) \right| = 0.$$

Choose in condition (iv) of the comparison theorem, $\delta = 2\alpha \leq 1$, α given by (2.9). Using the bound obtained in (3.20), we have

$$\sum_{i=1}^{n} E \int_R |x|^{2+2\alpha} |dp'_{ni}(x) - dq'_{ni}(x)|$$

$$\leq E\{(1 + C) \sum_{i=1}^{n} E'|U_{ni}|^{2(1+\alpha)}\}$$

$$= (1 + C)\, E\{\max_{1\leq i\leq n} |C_{ni}|^{2(1+\alpha)}\}\, O(n^{-\alpha})$$

$$\to 0, \text{ as } n \to +\infty$$

by condition (2.9), lemma (3.1) and the C_r-inequality.

Thus $\sum_{i=1}^{n} U_{ni}$ and $\sum_{i=1}^{n} V_{ni}$ are asymptotically equivalent in law. Next, to obtain the characteristic function of $V_n = \sum_{i=1}^{n} V_{ni}$,

$$f_n(u) = E \exp\{iu V_n\} = E E' \exp\{iu V_n\}$$
$$= E \exp(iu \mu_n - \frac{1}{2} u^2 \sigma^2(\underset{\sim}{C}_n)).$$

Let $W_n(\cdot, \cdot)$ be the joint distribution of $(\mu_n, \sigma_n(\underset{\sim}{C}_n))$.

We then have

$$(3.21) \qquad f_n(u) = \int \exp(iu \mu - \frac{1}{2} u^2 \sigma^2) \, d\, W_n(\mu, \sigma^2)$$

which by continuity theorem converges to a characteristic function $f_W(u)$ (say) if and only if $W_n(\cdot, \cdot)$ converges weakly.

It remains to show that the remainder terms all converge in probability unconditionally. Recall that conditional convergence has been established in Theorem 2.1. Thus the proof of Theorem 2.2 will be complete if we prove the following:

<u>Proposition 3.2</u>. <u>Let</u> $\{Q_n\}$ <u>be the family of conditional probabilities determined by</u> $\{\underset{\sim}{C}_n\}$. <u>Let this family be regular with respect to</u> $(\Omega, \mathcal{G})$. <u>Let the conditional probability</u> $P(|D_n| > \delta \,|\underset{\sim}{C}_n) \to 0$. <u>If</u> $\max_{1 \leq i \leq n} |C_{ni}| = O_p(1)$, <u>then</u> $P(|D_n| > \delta) \to 0$ <u>unconditionally</u>.

<u>Proof</u>. Let ν_n be the probability measure determined by $\underset{\sim}{C}_n$. Then

$$(3.20) \qquad P(|D_n| > \delta) = \int P(|D_n| > \delta \,|\underset{\sim}{C}_n) \, d\nu_n.$$

Now, given $\epsilon > 0$, there is a constant K such that $P(\max_{1 \leq i \leq n} |C_{ni}| > K) < \epsilon$.

Now

$$\int P(|D_n| > \delta|\underset{\sim}{C}_n) d\nu_n = \int_{[\max_{1 \leq i \leq n} |C_{ni}| \leq K]} P(|D_n| > \delta|\underset{\sim}{C}_n) d\nu_n$$

$$+ \int\limits_{[\max_{1\le i\le n} |C_{ni}| > K]} P(|D_n| > \delta | \underset{\sim}{C}_n)\, d\nu_n$$

$$< \int\limits_{[\max_{1\le i\le n} |C_{ni}| \le K]} P(|D_n| > \delta | \underset{\sim}{C}_n)\, d\nu_n + \epsilon .$$

Setting

$$(3.21) \qquad f_{n,\delta} = P(|D_n| > \delta | \underset{\sim}{C}_n)\, I_{[\max_{1\le i\le n} |C_{ni}| \le K]}$$

and noting that $\lim_{n\to\infty} f_{n,\delta} = 0$ a.s., it suffices to prove that

$$\lim_{n\to+\infty} \int f_{n,\delta}\, d\nu_n = 0.$$

Since $\{Q_n\}$ forms a family of regular, conditional probabilities in the product space, $A_n \in \mathcal{B}_1 \times \cdots \times \mathcal{B}_n$

$$\nu_n(A_n) = \int Q_1(d\omega_1) \int Q_2(\omega_1, d\omega_2) \cdots \int Q_n(\omega_1, \cdots, \omega_{n-1}; d\omega_n) I(A_n).$$

Then by a theorem of Ionescu-Tulcea (1949, 50), (see also Loève (1963), p. 137), there is a probability measure Q on $\mathcal{C}_0$, determined by $\{\nu_n\}$, and so by the generalized Fatou Lemma, $\lim_{n\to\infty} \int f_{n,\delta}\, d\nu_n \le \int \lim_{n\to\infty} \sup f_{n,\delta}\, dQ = 0$. The proof is complete.

Clearly, the proof of Theorem 2.3 is trivial. Proof of Corollary 2.1 follows along similar lines, by a direct application of Theorem A, page 381 of Loève (1963).

4. Some Extensions: We first obtain an extension of our results of Section 2 when the scores are given by (1.3).

Proposition 4.1: Let the hypotheses of Theorems 2.1 and 2.2 be satisfied. Let, in addition, φ be the inverse of a cdf. Let T_n be defined as before but with the scores now given by (1.3). Then the conclusions of Theorems 2.1, 2.2, and 2.3 as well those of Corollary 2.1 hold.

Proof. Denote $T_n^* = n^{-1/2} \sum_{i=1}^{n} C_{ni}\, a_n(R_{ni})$ with $a_n(i)$ given by (1.3). We have to prove

(4.1) $\quad T_n - T_n^* \to 0$ in probability.

Since $\max_{1 \leq i \leq n} |C_{ni}| = O_p(1)$, it suffices to prove that

(4.2) $\quad P(|T_n - T_n^*| > \delta | \underset{\sim}{C}_n) \to 0.$

Then Proposition 3.2 entails (4.1). But (4.2) is an immediate consdquence of Puri-Rajaram (1977), whenever $\max_{1 \leq i \leq n} |C_{ni}|$ is bounded, which is valid on a set of probability close to 1.

We now indicate a multivariate extension of some of our results. Usually extensions from the univariate to multivariate situations are fairly straightforward in view of the Cramér-Wold criterion. Since our limit laws can be nonnormal, it is not immediately obvious how Theorem 2.2 would extend in a multivariate setup. We limit ourselves to an extension of Theorem 2.3.

Let $\underset{\sim}{X}'_{ni} = (X_{ni}^{(1)}, \cdots, X_{ni}^{(p)})$, $n \geq 1$, $i = 1, \cdots, n$ be a sequence of p-variate random vectors ($p \geq 1$), with continuous cdf.'s $F_{ni}(\underset{\sim}{x})$, $1 \leq i \leq n$, $\underset{\sim}{x} \in R^p$. Let $\{C_{ni}: 1 \leq i \leq n\}$ be the sequence of stochastic predictor variables. As before we assume the vectors $\underset{\sim}{X}_{ni}$ to be conditionally independent, given $(\underset{\sim}{C}_n)$. Consider now the statistics $\underset{\sim}{T}'_n = (T_n^{(1)}, \cdots, T_n^{(p)})$,

$T_n^{(\nu)} = n^{-1/2} \sum_{i=1}^{n} c_{ni}\, a_n^{(\nu)} (R_{ni}^{(\nu)})$ where the $a_n^{(\nu)}(i)$'s are given as in Section 1, and satisfy the assumptions of Theorem 2.1. Clearly, each $T_n^{(\nu)} - \mu_n^{(\nu)}$ can be expressed as a partial projection on $\underset{\sim}{X}_n^{(\nu)}$ and $\underset{\sim}{C}_n$ by Theorem 2.1. Let the conditional variance-covariance matrix of $(\psi_n^{(1)} (\underset{\sim}{X}_n^{(1)}, \underset{\sim}{C}_n), \cdots, \psi_n^{(p)} (\underset{\sim}{X}_n^{(p)}, \underset{\sim}{C}_n))$ where

$\psi_n^{(\nu)} (\underset{\sim}{X}_n^{(\nu)}, \underset{\sim}{C}_n) = n^{-1/2} \sum_{i=1}^{n} \psi_{ni}^{(\nu)} (X_{ni}^{(\nu)}, \underset{\sim}{C}_n)$ be given by $\Sigma_n(\underset{\sim}{C}_n)$. Let $\mu_n^{(\nu)}, \psi_n^{(\nu)} (\cdot, \cdot)$ be defined in the usual manner (cf. Theorem 2.1). Then we have

Theorem 4.1. Let φ_ν and $\underset{\sim}{C}_n$ satisfy conditions of Theorem 2.1 and let $\lim_{n\to\infty} \Sigma_n (\underset{\sim}{C}_n) = \Sigma_0$ where Σ_0 is a positive definite (nonrandom) matrix. Then $(\underset{\sim}{T}_n - \underset{\sim}{\mu}_n)$ is asymptotically normal with mean $\underset{\sim}{0}$ and covariance matrix Σ_0.

The proof follows by using Theorem 2.1, Cramér-Wold criterion, extended c_r-inequality and Lemma 2.1.

REFERENCES

[1] Hájek, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist., 33, 1124-1147.

[2] Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39, 325-346.

[3] Ionescu-Tulcea (1949,50). Measures dans les espaces produits. Atti. Accad. Naz. Lincei. Rend., 7.

[4] Loève, M. (1963). Probability Theory, 3rd ed. Van Nostrand, N.Y.

[5] Puri, M. L. and Rajaram, N. S. (1977). Asymptotic multinormality and remainder terms of linear rank vectors under alternatives. J. Statist. Planning and Inference (submitted).

[6] Puri, M. L. andSen, P. K. (1971). Nonparametric Methods in Multivariate Analysis. John Wiley, N.Y.

Department of Mathematics
Swain Hall East
Indiana University
Bloomington, Indiana 47401

Department of Mathematics
Merrill Hall
Kent State University
Kent, Ohio 44242

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 **REPORT DOCUMENTATION PAGE** — READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR TR-78-0433	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): 6 LIMIT THEOREMS RELATED TO DISTRIBUTIONS OF CERTAIN LINEAR RANK STATISTICS WITH STOCHASTIC PREDICTORS.		5. TYPE OF REPORT & PERIOD COVERED: 9 Interim rept.,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 Madan L. Puri [redacted] Navaratna S. Rajaram		8. CONTRACT OR GRANT NUMBER(s): 15 AFOSR-76-2927
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Indiana University, Department of Mathematics, Bloomington, Indiana 47401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 16 17 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE: February 28, 1978
		13. NUMBER OF PAGES: 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 11 28 Feb 78 12 20p.		15. SECURITY CLASS. (of this report): UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear rank statistics, stochastic linear rank statistics, stochastic predictors, weighted normal distribution.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $\{X_{ni},\ 1 \le i \le n,\ n \ge 1\}$ be independent random variables with respective continuous cumulative distribution functions $\{F_{ni},\ 1 \le i \le n,\ n \ge 1\}$ and let $\{R_{ni},\ 1 \le i \le n,\ n \ge 1\}$ be the corresponding ranks. Consider a stochastic linear rank statistic

continued--

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

402 522 JOB

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. ABSTRACT (cont'd)

$T_n = n^{-\frac{1}{2}} \sum_{i=1}^{n} c_{ni} \varphi\left(\frac{R_{ni}}{n+1}\right)$ where $(c_{n1}, \ldots, c_{nn})$ are random variables, not necessarily independent, and φ a finite function on (0,1). Under suitable assumptions, it is shown that the limit distribution of T_n is weighted normal which degenerates to the normal in extreme cases.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)